NRC Research Press

ARTICLE

Modeling height-diameter curves for prediction

Lauri Mehtätalo, Sergio de-Miguel, and Timothy G. Gregoire

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

Abstract: Individual tree heights are needed in many situations, including estimation of tree volume, dominant height, and simulation of tree growth. However, height measurements are tedious compared to tree diameter measurements, and therefore height–diameter (H–D) models are commonly used for prediction of tree height. Previous studies have fitted H–D models using approaches that include plot-specific predictors in the models and those that do not include them. In both these approaches, aggregation of the observations to sample plots has usually been taken into account through random effects, but this has not always been done. In this paper, we discuss four alternative model formulations and report an extensive comparison of 16 nonlinear functions in this context using a total of 28 datasets. The datasets represent a wide range of tree species, regions, and ecological zones, consisting of about 126 000 measured trees from 3717 sample plots. Specific *R*-functions for model fitting and prediction were developed to enable such an extensive model fitting and comparison. Suggestions on model selection, model fitting procedures, and prediction are given and interpretation of the predictions from different models are discussed. No uniformly best function, model formulation, or model fitting procedure was found. However, a 2-parameter Näslund and Curtis function provided satisfactory fit in most datasets for the plot-specific H–D relationship. Model fitting and height imputation procedures developed for this study are provided in an *R*-package for later use.

Key words: mixed-effects, fixed-effects, plot-specific, marginal, prediction, lmfor, imputation, height-diameter, forest inventory.

Résumé : La hauteur des arbres individuels est nécessaire dans de nombreuses situations, y compris pour l'estimation du volume des arbres, de la hauteur dominante et pour la simulation de la croissance des arbres. Cependant, la mesure de la hauteur est une tâche fastidieuse comparativement à la mesure du diamètre des arbres et les modèles hauteur-diamètre (H–D) sont par conséquent couramment utilisés pour prédire la hauteur des arbres. Des études antérieures ont ajusté des modèles H–D en utilisant des approches qui incluent des prédicteurs spécifiques aux parcelles-échantillons dans les modèles et des approches qui n'en comprennent pas. Dans ces deux approches, l'agrégation des observations au sein des parcelles-échantillons a généralement été prise en compte par des effets aléatoires, mais cela n'a pas toujours été le cas. Dans cet article, nous discutons des quatre formulations possibles de modèles qui en résultent et présentons une comparaison exhaustive de 16 fonctions non linéaires dans ce contexte en utilisant un total de 28 ensembles de données. Ces ensembles de données, composés d'environ 126 000 arbres mesurés à partir de 3717 parcelles-échantillons, représentent un large éventail d'espèces d'arbres, de régions et de zones écologiques. Des fonctions spécifiques à l'ajustement des modèles et à la prédiction ont été développées dans le logiciel R pour permettre une aussi vaste étude d'ajustement et de comparaison des modèles. Nous donnons des suggestions sur la sélection des modèles, les procédures d'ajustement des modèles et sur la prédiction. Nous discutons de l'interprétation des prédictions des différents modèles. Aucune fonction, formulation de modèle ni méthode d'ajustement ne s'est révélée régulièrement meilleure. Cependant, une fonction de Näslund et Curtis à deux paramètres a fourni un ajustement satisfaisant dans la plupart des ensembles de données pour la relation H–D spécifique aux parcelles-échantillons. Les procédures d'ajustement des modèles et d'imputation des hauteurs élaborées pour cette étude sont fournies dans une extension du logiciel R pour une utilisation ultérieure. [Traduit par la Rédaction]

Mots-clés : effets mixtes, effets fixes, échelle de la placette-échantillon, prédiction, extension lmfor, imputation, relation hauteur-diamètre, inventaire forestier.

Introduction

Information on tree heights is essential in forest inventories for computing tree volumes. Also, growth and yield simulators usually need information on tree height, either at the individual tree, plot, or stand level, to predict forest dynamics, dominant height, and site index (e.g., Huang et al. 2000). However, field measurement of tree height is rather tedious compared to measuring tree diameter. That is why many forest inventories save time and effort by predicting tree heights using height–diameter (H–D) models instead of direct measurements. Height measurements from a subsample of trees on each sample plot or sample plot cluster may be utilized for improved prediction of the local H–D curve.

Two kinds of H–D models have been reported in the literature: models that express the height as a function of tree diameter only and models that include additional stand-level predictors in the model. Soares and Tomé (2002) call these two model types local and regional H–D models, respectively, whereas others use the term "generalized" for the latter model type (Temesgen and von Gadow 2004; Paulo et al. 2011). However, both these models can be fitted at the plot-level and regionally, as we will discuss later on. We therefore use the term *simple* for the model without

Received 9 February 2015. Accepted 8 April 2015.

L. Mehtätalo. University of Eastern Finland, School of Computing, P.O. Box 111, FI-80101 Joensuu, Finland.
S. de-Miguel. University of Eastern Finland, School of Forest Sciences, P.O. Box 111, FI-80101 Joensuu, Finland; Departament de Producció Vegetal i Ciència Forestal, Universitat de Lleida-Agrotecnio Center (UdL-Agrotecnio), Av. Rovira Roure, 191, E-25198 Lleida, Spain.
T.G. Gregoire. School of Forestry and Environmental Studies, Yale University, New Haven, CT 06511, USA.
Corresponding author: Lauri Mehtätalo (e-mail: lauri.mehtatalo@uef.fi).

 Published at www.nrcresearchpress.com/cjfr on 9 April 2015.

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

stand-level predictors in contrast to the *generalized* model that also includes stand-level predictors. An important special case of the latter model type is a model with any plot-specific aggregate of tally trees as an additional predictor (e.g., quadratic mean, arithmetic mean, or the basal-area median diameter, stand density (trees per ha) or basal area) (e.g., Siipilehto 1999). Indeed, such a model does not need additional measurements other than tree diameters of the plot and is therefore as widely applicable for prediction as the simple model.

H–D models often aim at predictions at the plot level. Therefore, the models should express the plot-specific H–D relationship. The generalized models take a step towards this direction by including plot-specific predictors. However, they still assume that plots with similar values of stand-level predictors also have similar H–D curves. Therefore, the solution is not fully satisfactory. An approach relying purely on plot-specific relationships is provided by allowing plot-specific parameters in the models. A possible way to do that is to fit separate models by plot. However, this approach is usually inappropriate in practice, since it requires a large number of trees per plot and does not allow prediction for plots outside the modeling data. The data needs can be reduced by using plot index as a categorical predictor in the model, but the problem of prediction for new plots still remains. In addition, such a model does not allow any other plot-specific characteristics (e.g., mean diameter) as predictors, because it leads to over-parameterization. An approach that also enables prediction for new plots is provided by the mixed-effects models, where a fixed part provides the predicted H–D curve for a typical sample plot of the modelling data (fixed-effect prediction). A prediction augmented with predicted values of the random effects (random-effect prediction) provides a calibrated prediction, which realistically describes the plot-specific H–D relationship. Especially in situations where sample trees are available from the plot in question, the mixed-effects models have proven highly useful, since they allow for developing plot-specific H–D curves with minimal sample size per plot (e.g., Lappi 1991, 1997; Mehtätalo 2004, 2005; Calama and Montero 2004; Paulo et al. 2011; Sirkiä et al. 2014).

Some authors have criticized the fixed-effect prediction of a mixed-effects model for poor predictive performance when compared to a prediction from fixed-effect models using RMSE and empirical bias as the criteria (e.g., Temesgen et al. 2008). Therefore, they have fitted H–D models as fixed-effects models, ignoring the grouping of the observations into sample plots. Most of these articles have used a generalized model (e.g., Arabazis and Burkhart 1992; Houghton and Gregoire 1993; Lynch et al. 2005), but there are also published simple regional models (e.g., Moore et al. 1996; Huang et al. 2000). These models are justified if the aim is *marginal prediction*, i.e., prediction of the mean height of trees with a given diameter (and given plot-specific predictors in the case of generalized model) in the region of interest. However, they do not provide *a plot-specific* prediction for a plot of interest or for a typical plot of given characteristics. In addition, the fixed-effects models are not satisfactory from the viewpoint of statistical inference, since they do not yield reliable estimates of parameter uncertainty.

One could also classify the H–D models as nonlinear and linear. The linear models are traditionally preferred because of the well-established theory of estimation and inference. However, fitting nonlinear mixed-effects model is nowadays seldom problematic, nor is inference a major issue in most forest inventory applications. In addition, articles comparing different functions for the H–D relationship (e.g., Huang et al. 1992; Arabazis and Burkhart 1992) have found nonlinear models better than the linear models. Even though some nonlinear models could be easily linearized, we focus on such nonlinear models that have fitted well in literature and where the response is untransformed tree height. Our main reason for using nonlinear models is to avoid the need for back-transformation bias correction. However, one should remember that there may still be some bias problem, because parameter estimates of nonlinear models, and therefore also the predictions, are only asymptotically unbiased (Box 1971).

In this study, we explored the fit of different nonlinear functions for the H–D relationship in 28 different datasets, representing a wide range of regions, ecological zones, and tree species. To demonstrate the differences between the marginal and plot-specific H–D relationships, the fitting was done both using the fixed-effects and mixed-effects modeling approaches. In addition, we developed generalized mixed-effects models for four example datasets to demonstrate that the best-fitting models can differ largely in different datasets and to demonstrate the predictive properties of generalized and simple mixed- and fixed-effects models.

Material

The empirical data of this study includes 19 different growth and yield experiments and inventory datasets from Europe, Asia, North America and South America, representing ecological zones from tropical to boreal conditions. Five of the datasets included several measurement occasions, which were treated as a separate datasets. The maximum number of measurement occasions per experiment was, however, restricted to three. This yielded a total of 28 datasets, with a total of 126 019 trees from 3717 sample plots. More details are given in Table 1.

Methods

The simple mixed-effects model

Denote the height of tree j on plot i by h_{ij} and the corresponding tree diameter at breast height by d_{ij}. Depending on the applications, i could be the index of a forest stand or region as well, but in this paper we will call it the plot index for clarity of presentation. The mixed-effects model for plot i is defined as,

$$(1) \qquad h_{ij} = f(d_{ij}; \boldsymbol{\beta}_i) + e_{ij}$$

where $f(d_{ij}; \boldsymbol{\beta}_i)$ is the systematic part of the model, and e_{ij} is the unexplained residual error. If the model form is correct and parameters are known, $f(d_{ij}; \boldsymbol{\beta}_i)$ provides a conditional mean $E(h_{ij}|d_{ij})$ on the plot. The systematic part can be either linear or nonlinear with respect to $\boldsymbol{\beta}_i$; our notation covers both options. However, the most widely used functions are nonlinear unless some transformation is made to h_{ij}. The p-element vector $\boldsymbol{\beta}_i$ includes parameters of the mean function for plot i, with $p = 2$ or $p = 3$ parameters.

The systematic part is allowed to vary between sample plots through inclusion of random effects. For the *simple mixed-effects model*, we define,

$$(2) \qquad \boldsymbol{\beta}_i = \boldsymbol{B} + \boldsymbol{b}_i$$

where $\boldsymbol{B}$ represents the parameters for a typical plot in the whole population of plots, and plot effects $\boldsymbol{b}_i$ express the difference in the parameters of plot i from the typical plot. The plot effects are assumed to have a common multivariate normal distribution with mean 0 and variance–covariance matrix var($\boldsymbol{b}_i$) = $\boldsymbol{D}$ for all values of i.

The residual errors are assumed to be independent, with zero mean and common constant variance var(e_{ij}) = σ^2. For heteroscedastic data, one can also model the variance parametrically. For example, one can assume the power-type variance function,

$$(3) \qquad \mathrm{var}(e_{ij}) = \sigma^2 \|w_{ij}\|^{2\delta}$$

which allows both increasing ($\delta > 0$) or decreasing ($\delta < 0$) variance with respect to the predictor w_{ij}. The value $\delta = 0.5$ implies a linear

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

Table 1. Summary of the modeling datasets.

Data set	Latin name	Country	N	K	$\overline{n_i}$	d_{min}	$\overline{d}$	d_{max}	h_{min}	$\overline{h}$	h_{max}
Scots pine A	*Pinus Sylvestris*	Finland	4234	103	41	1.5	14.5	51.0	1.4	13.2	35.1
Norway spruce A	*Picea abies*	Finland	2513	51	49	2.9	17.2	57.0	2.1	13.7	29.8
Scots pine B	*Pinus Sylvestris*	Finland	1644	66	25	3.0	20.0	49.1	1.6	17.3	33.1
Norway spruce B	*Picea abies*	Finland	3020	66	46	0.9	11.5	52.3	1.4	9.9	33.2
Birch A	*Betula pendula. B. pubescens*	Finland	1673	72	23	1.6	8.7	48.8	1.8	10.0	29.8
Norway spruce C	*Picea abies*	Finland	1252	31	40	5.0	14.3	68.8	1.5	12.9	34.3
Turkish red pine	*Pinus brutia*	Syria. Lebanon	1283	114	11	5.0	27.3	96.9	3.5	13.7	35.1
Aleppo pine	*Pinus halepensis*	Spain	16378	1016	16	7.5	32.6	174.0	2.0	14.6	41.0
Canarian island pine	*Pinus canariensis*	Spain	7327	870	8	7.5	19.5	74.8	2.0	7.7	23.0
Loblolly pine 1	*Pinus taeda*	VA, USA	5634	99	57	1.3	13.9	34.3	1.5	10.9	23.8
Loblolly pine 2	*Pinus taeda*	VA, USA	4895	99	49	3.3	18.2	37.6	4.6	15.8	26.8
Loblolly pine 3	*Pinus taeda*	VA, USA	4171	99	42	5.1	20.8	42.9	4.9	18.8	31.4
Lodgepole pine 1	*Pinus contorta*	BC, Canada	10817	140	77	0.1	6.0	25.5	1.3	7.4	21.3
Lodgepole pine 2	*Pinus contorta*	BC, Canada	9336	141	66	0.3	8.9	31.0	1.3	8.7	22.8
Lodgepole pine 3	*Pinus contorta*	BC, Canada	5903	93	63	0.7	12.6	29.8	1.4	12.4	24.2
Eucalyptus clone	*Eucalyptus urograndis*	Brazil	1141	191	6	6.2	19.4	34.1	12.0	30.0	41.0
Blue gum A	*Eucalyptus globulus*	Bolivia	6554	50	131	0.1	3.8	18.2	1.4	6.0	19.1
Blue gum B1	*Eucalyptus globulus*	Bolivia	884	6	147	1.0	9.5	31.7	1.9	9.6	28.5
Blue gum B2	*Eucalyptus globulus*	Bolivia	1261	6	210	1.0	10.1	33.7	1.7	11.1	30.0
Centrolobium 1	*Centrolobium tomentosum*	Bolivia	2199	46	48	1.2	11.2	28.3	1.8	11.1	20.5
Centrolobium 2	*Centrolobium tomentosum*	Bolivia	2167	46	47	1.2	12.6	30.3	2.2	12.6	22.1
Centrolobium 3	*Centrolobium tomentosum*	Bolivia	2023	44	46	2.5	13.4	31.3	2.2	13.6	25.8
Brasilian firetree	*Schizolobium parahyba*	Bolivia	2631	46	57	0.8	8.8	33.2	1.4	8.7	27.0
Teak 1	*Tectona grandis*	Bolivia	4928	62	79	1.0	6.6	41.5	1.4	6.5	29.6
Teak 2	*Tectona grandis*	Bolivia	3444	43	80	1.0	8.1	44.3	1.4	7.9	29.5
Mixed tropical	Multi-species	Bolivia	15049	41	367	8.2	23.7	115.6	2.5	12.6	36.9
Balsa 1	*Ochroma pyramidale*	Bolivia	2943	53	56	0.8	8.7	19.8	1.5	8.6	21.7
Balsa 2	*Ochroma pyramidale*	Bolivia	715	23	31	1.5	9.9	22.7	1.4	9.8	17.5

Note: Whenever two datasets of same species have been used, a capital letter is used to denote different independent datasets and an Arabic number to denote different measurement occasions of the same dataset. N: the number of trees; K: the number of sample plots; $\overline{n_i}$: mean number of trees per plot; d_{min}, $\overline{d}$, d_{max}: the minimum, mean and maximum diameter, cm; h_{min}, $\overline{h}$, h_{max}: the minimum, mean and maximum height, m.

relationship between the variance and tree diameter, and $\delta = 1$ implies a linear relationship between standard deviation and the predictor. A commonly used formulation of the variance function is the "power-of-the-mean" type formulation $w_{ij} = \hat{y}_{ij}$, which requires an iterative estimation procedure. In this paper, we consider two alternative predictors in the variance function: tree diameter $w_{ij} = d_{ij}$ and truncated relative tree diameter $w_{ij} = \max(1, z_{ij} + 3)$, where

$$(4) \qquad z_{ij} = \frac{d_{ij} - m_i}{s_i}$$

and m_i and s_i are the mean and standard deviation of tree diameter on plot i. Adding the constant 3 to the standardized diameter and thereafter truncating at the value 1 make the predictor always positive. Furthermore, because most trees (approximately 97.5% of them) should have relative diameter above –2, this parameterization provides a nice interpretation of σ^2 as the (constant) error variance for the smallest trees of the plot; for the rest of the trees, the variance may either increase or decrease, depending on the value of parameter δ.

The fitting of a mixed-effects model involves estimation of model parameters $\boldsymbol{B}$, $\boldsymbol{D}$, σ, and possibly δ. In addition, the plot-effects $\boldsymbol{b}_i$ are obtained as a by-product from the estimation of a nonlinear model. The standard method of estimation is based on restricted maximum likelihood, and tests on nested models are based on the likelihood ratio (random part) and conditional F tests (fixed part). Non-nested models can be compared using Akaike and Bayesian information criteria (AIC, BIC) (Pinheiro and Bates 2000).

The generalized mixed-effects model

In the *generalized mixed-effects model*, parameter vector $\boldsymbol{\beta}_i$ of model (1) is written as a function of plot-specific predictors $\boldsymbol{x}_i$ and random effects $\boldsymbol{b}_i$ as follows,

$$(5) \qquad \boldsymbol{\beta}_i = \boldsymbol{B}(\boldsymbol{x}_i; \boldsymbol{\gamma}) + \boldsymbol{b}_i$$

The vector $\boldsymbol{\gamma}$ is a vector of parameters that are estimated in model fitting. The function $\boldsymbol{B}(\boldsymbol{x}_i; \boldsymbol{\gamma})$ is assumed to be of the linear form $\boldsymbol{x}_i'\boldsymbol{\gamma}$. The assumptions on the residual errors and random effects are the same as earlier for the simple models.

Some stand-level predictors $\boldsymbol{x}_i$ can be computed directly using the sample tree diameters. Such predictors are, for example, the means and variances of the diameters or basal area or stand density per area unit, if the sampling design allows such computation using the available diameter sample. A model including these predictors does not require additional information compared to the simple mixed-effects model. In observational H–D datasets, these stand-level aggregates may be correlated with the random effects of the simple mixed-effects model, meaning that the random effects do not have a common distribution if these aggregates are not included as fixed predictors in the model. Such a model may be inappropriate both for inference and for prediction. However, reformulating the fixed part by including the correlated aggregates (with appropriate transformations) solves the problem.

Plot-specific prediction using mixed-effects models

In the mixed-effects model (eqs. (1) and (2) or (5)), prediction is possible at two levels of grouping. A prediction called *fixed-effect prediction* is obtained by evaluating the random effects at their expected value, i.e., when $\boldsymbol{b}_i = \boldsymbol{0}$. The resulting function provides the plot-specific H–D curve for a typical plot of the modeling

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

dataset. Such predictions may be useful for a sample plot that does not have any height sample trees available for prediction of the plot effects, but one wants to produce a realistic plot-specific H–D curve. The fixed-effect prediction can also be interpreted as a marginal (population-averaged) prediction in some special situations.

Improved plot-specific prediction is provided by the *random-effect prediction*, where the predictions of random effects for the plot in question are utilized. Details on the procedures of random-effect prediction are not provided here, but readers are referred to Pinheiro and Bates (2000) for the case where prediction is done during the model fitting stage; please refer to Meng and Huang (2009); Calama and Montero (2004) and Sirkiä et al. (2014) for the case where prediction is done afterwards (mixed model calibration). These resulting curves are very similar to the curves that would result from separately fitting each plot, especially if the number of trees per plot is high. Nevertheless, by pooling observations from all plots together and stipulating random plot effects, far fewer observations per plot are needed to enable prediction. Even one measured sample tree per plot is enough for prediction of $\boldsymbol{b}_i$, although increasing the number of sample trees improves the accuracy of the resulting prediction. For mixed model calibration of linear models, see Lappi (1986); Lappi and Bailey (1988); Lappi (1997) and Mehtätalo (2004).

Simple and generalized fixed-effects models for marginal prediction

Instead of the H–D relationship at the plot level, one might be interested in the mean height of trees with given diameter and plot-specific predictors in the population of trees, *the marginal H–D relationship*. This relationship is also commonly called the population-averaged prediction. We consider two kinds of marginal prediction: (*i*) the mean height for a given diameter over whole population of trees (simple marginal prediction) and (*ii*) the mean height for a given diameter in the subpopulation of trees that originates from sample plots with similar plot-specific characteristics (generalized marginal prediction).

A practical and straightforward method to estimate the marginal relationship is to fit an ordinary or weighted least-squares regression by ignoring the hierarchy of the data in model fitting. For simple marginal prediction, this means fitting *the simple fixed-effects model*,

$$h_{ij} = f(d_{ij}; \boldsymbol{\beta}) + e_{ij} \tag{6}$$

where the parameter $\boldsymbol{\beta}$ is common for all plots. The generalized marginal predictions are provided by the *generalized fixed-effects model*, where $\boldsymbol{\beta}$ is replaced by a (usually linear) function of plot-specific predictors, $\boldsymbol{x}_i'\boldsymbol{\gamma}$, the parameters $\boldsymbol{\gamma}$ of which are common to all plots.

The fixed-effects model assumes all observations to be independent realizations from a common population model. This does not hold because of grouping of trees into sample plots, which results in dependence among observations, which might be caused, inter alia, from stand or microsite conditions that affect growth. From this point of view, the fixed-effects model is inaptly formulated, and the related tests and confidence intervals cannot be trusted. In addition, the estimators of the parameters are not fully efficient. Fitting of marginal models is commonly justified from the predictive point of view (Temesgen et al. 2008; Garber et al. 2009), where the fixed-effect prediction of a mixed-effects model may have clearly higher RMSE than a prediction of a fixed-effects model. However, straightforward use of the RMSE as the evaluation criteria leads to favoring fixed-effects models estimated using OLS, because OLS estimation uses MSE as the model fitting criteria.

A better approach for marginal prediction would be to use the mixed-effects model. For generalized marginal prediction, this means averaging the prediction from a generalized mixed-effects models over the distribution of random effects (de-Miguel et al. 2012). For the simple marginal prediction, one might be tempted to use a similar approach with simple mixed-effects model. However, this does not provide a satisfactory solution in the common situation where the simple mixed-effects model is improperly formulated because random effects correlate with plot-specific aggregates of diameter (e.g., mean diameter and stand density). A properly formulated generalized mixed-effects model should be used for simple marginal prediction in this case, by averaging the generalized marginal prediction over the distribution of the plot-specific predictors for the given tree diameter. For this average, a joint distribution of the model predictors in the tree population of interest is needed.

Modeling the H–D relationship in different datasets

The applied nonlinear functions

A set of 16 functions was selected from previously published studies on H–D modeling (Table 2). To avoid problems with back transformation bias, we used only functions that expressed tree height without transformations. All functions were nonlinear, with either 2 (7 functions) or 3 parameters (9 functions). All functions were parameterized so that parameter *a* defines the scale of the H–D model, which is the major dimension of variability among sample plots.

Simple models for all datasets

We fitted each function to each dataset as a simple mixed-effects model (eq. (1)) using package nlme of *R*-environment (R Development Core Team 2010; Pinheiro and Bates 2000). All parameters were initially assumed random, with a general positive definite variance–covariance structure for the random effects. If this model did not converge, then the number of random parameters was reduced (by dropping random effects associated with parameter *c*, *b* in this order) until the model did converge. However, if model fitting with even one random parameter (scale parameter *a*) was not successful, the modeling effort was abandoned for the data in question, and the model was classified as non-converging.

To illustrate the differences between the marginal and plot-specific H–D relationship, we fitted also simple fixed-effects models (eq. (6)) to each dataset, regardless of the problems related to efficiency and inference on the models. These models were fitted to the datasets by using function gnls of the nlme package.

The simple models (both fixed-effects and mixed-effects models) were fitted by assuming constant error variance, which leads to a minimum RMSE when the number of trees per plot is large. Therefore, the use of RMSE is justified at this stage in comparing the fit of functions with a similar number of parameters in the same dataset. For each dataset and model, the models were ranked according to the RMSE of height prediction in the model fitting dataset; the evaluated plot-specific prediction included the predicted random effects. The ranking was done separately for 2- and 3-parameter models and for mixed-effects and fixed-effects models. For each of these classes, the number of first ranks, ranking among the best three models, mean rank, and number of datasets with non-convergence were computed for each model.

The best-fitting functions were re-estimated by using the power-type variance function, using tree diameter as the predictor in the function. Models with and without variance functions were compared using the LR test of REML fits, as implemented in the nlme package. In addition, the best-fitting 2- and 3-parameter mixed-effects models were compared to each other by using the Akaike and Bayesian information criteria (AIC and BIC) using ML fits, which are the only possible criteria to compare fits of non-nested models.

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

Table 2. The applied H-D functions.

Number	Function name	Equation	References
2-parameter functions			
1	Näslund	$H(D) = BH + \frac{D^2}{(aD + b)^2}$	Näslund (1937), Peschel (1938)
2	Curtis	$H(D) = BH + \frac{aD}{(1 + D)^b}$	Curtis (1967)
3	Schumacher	$H(D) = BH + a\exp(-bD^{-1})$	Schumacher (1939), Michailoff (1943), Curtis (1967)
4	Meyer	$H(D) = BH + a(1 - \exp(-bD))$	Meyer (1940), Curtis (1967)
5	Power	$H(D) = BH + aD^b$	Stoffels and van Soest (1953)
6	Michaelis–Menten	$H(D) = BH + aD/(b + D)$	Menten and Michaelis (1913), Huang et al. (1992)
7	Wykoff	$H(D) = BH + \exp(a - b(D + 1)^{-1})$	Wykoff et al. (1982)
3-parameter functions			
8	Prodan	$H(D) = BH + \frac{D^2}{aD^2 + bD + c}$	Strand (1959)
9	Logistic	$H(D) = BH + \frac{a}{1 + b\exp(-cD)}$	Pearl and Reed (1920), Huang et al. (1992)
10	Chapman-Richards	$H(D) = BH + a(1 - \exp(-bD))^c$	Richards (1959), Huang et al. (1992)
11	Weibull	$H(D) = BH + a(1 - \exp(-bD^c))$	Weibull (1951), Huang et al. (1992)
12	Gomperz	$H(D) = BH + a\exp(-b\exp(-cD))$	Gomperz (1825), Huang et al. (1992)
13	Sibbesen	$H(D) = BH + aD^{bD^{-c}}$	Sibbesen (1981), Huang et al. (1992)
14	Korf	$H(D) = BH + a\exp(-bD^{-c})$	Lundqvist (1957), Flewelling and de Jong (1994)
15	Ratkowsky	$H(D) = BH + a\exp\left(\frac{-b}{D + c}\right)$	Ratkowsky (1990), Huang et al. (1992)
16	Hossfeld IV	$H(D) = BH + \frac{a}{1 + \frac{1}{bD^c}}$	Peschel (1938)

Note: The references give the original reference and the first use in H–D modeling. Naming follows Zeide (1993) when applicable. H = tree height, D = tree diameter at breast height, BH = breast height, a, b, c = parameters of the equation.

Because of a large number of datasets and models, functions for model fitting, evaluation, and tree height prediction were implemented in the *R*-package lmfor (Mehtätalo 2015). These functions use the fitting routines of the nlme package inside the implemented functions but use reasonable default arguments and starting values for the parameters, thus being easy and fast to use in the context of height modeling and imputation. The initial values of the parameters were based on an OLS fit of the linearized forms of the models.

Generalized models for selected datasets

Detailed modeling of the H–D relationship was conducted for four different datasets. The datasets were selected so that they all had a different model selected as the best one in the simple model fitting stage. The generalized mixed-effects model fitting included a more detailed assessment of fit using plots of residuals and random effects, information criteria, conditional *F*-tests of fixed effects, and likelihood ratio tests on random effects and error variance. We assumed that no information other than the measured tree diameters and heights were available on the sample plots, therefore only the mean diameter and number of trees per hectare were evaluated as predictors in the models.

A generalized fixed-effects model was estimated for each dataset by fitting the final generalized mixed-effects model without random effects.

Results

Evaluation of simple models

In the mixed-effects model fitting, the best 2-parameter models were the functions called Näslund and Curtis (Table 3). Näslund's function was the best when ranked according to number of first ranks, whereas the Curtis model was the best when ranked according to number of ranks 1–3 or mean rank. Schumacher's function was the third best among the 2-parameter models. In the simple fixed-effects model, Näslund's model fitted the best according to all these criteria, followed by the models of Curtis and Meyer. Only one fit did not converge, and all estimated variance–covaraince matrices were positive definite. On the other hand, none of the fixed-effects model fits converged to all datasets.

From among the 3-parameter models, the best simple mixed-effects models were the models named Prodan, Gomperz, and Ratkowsky (Table 4). For the simple fixed-effects model, the best functions were the Chapman–Richards and Hossfeld IV functions. Convergence was a problem when fitting simple fixed-effects models. In fact, none of the models would converge to a solution when fitting the Aleppo pine data. Also, simple mixed-effects models had convergence problems with some datasets. A non-positive definite estimated variance–covariance matrix of random effects was obtained in 5–8 datasets for all models.

The variance function improved the fit in all but one dataset. However, the additional parameter introduced by the variance function increased the number of convergence problems compared to the fits without variance function.

Table 5 shows the best-fitting model for each dataset. Prodan's function provides Näslund's function as a special case, when $b = 2ac$, and Prodan's function was the best-fitting 3-parameter model for four datasets where Näslund's function was the best 2-parameter function. Otherwise, we did not notice any logical pattern in the fit of different models. For example, our datasets represented several different pine tree species, but the best fitting model was different for different pine species. The best-fitting models for the same dataset at different measurement occasions also varied without an obvious pattern.

The BIC suggested the more parsimonious 2-parameter model for 15 datasets, and the less conservative AIC suggested it for 9 datasets out of the 28. Three-parameter models were generally suggested for datasets with the largest number of observations, whereas 2-parameter models were sufficient for most smallest datasets. When considering the RMSE of height prediction, the 3-parameter model usually provided a very small reduction compared to the 2-parameter models.

Table 3. Evaluation of the simple two-parameter models according to the four criteria.

	Criteria	Näslund	Curtis	Schumacher	Meyer	Power	Mic.-Ment.	Wykoff
Mixed-effects	1st ranks	**12**	**7**	**8**	0	1	0	1
	Ranks 1–3	15	**26**	**20**	5	2	3	**16**
	Mean rank (sd)	**2.7 (1.5)**	**2.3 (1.2)**	**3.0 (1.9)**	4.7 (1.4)	5.9 (1.5)	5.8 (1.4)	3.6 (1.3)
	Conv. Prob's	**0**	**0**	**0**	**0**	**0**	1	**0**
Fixed-effects	1st ranks	**13**	**5**	0	**6**	2	3	0
	Ranks 1–3	**23**	**19**	8	**12**	6	**10**	8
	Mean rank (sd)	**1.9 (1.1)**	**2.8 (1.3)**	4.2 (1.5)	**2.5 (1.5)**	4.3 (2.0)	2.5 (1.5)	4.1 (1.1)
	Conv. Prob's	**4**	**3**	**5**	12	11	16	**3**

Note: The criteria are: 1st ranks is the number of first ranks among the datasets; ranks 1–3 gives the number of rankings among three best models; mean rank gives the mean rank of the model (the number in parentheses is the standard deviation of the ranks; Conv. Prob's gives the number of unsuccessful fits. The three best models according to each criteria are highlighted.

Table 4. Evaluation of the simple three-parameter models according to the four criteria.

	Criteria	Prodan	Logistic	Ch-Ri	Weibull	Gomperz	Sibbesen	Korf	Ratkowsky	Hossf. IV
Mixed-effects	1st ranks	**11**	**6**	1	0	4	0	0	**6**	0
	Ranks 1–3	**18**	12	10	8	**18**	0	2	**14**	2
	Mean rank (sd)	**3.0 (2.3)**	4.0 (2.2)	4.3 (1.8)	4.8 (1.7)	**2.8 (1.4)**	8.1 (1.3)	7.3 (1.7)	**3.6 (2.0)**	6.1 (1.3)
	Conv. Prob's	**0**	**0**	1	**0**	**0**	18	6	**0**	1
Fixed-effects	1st ranks	**3**	**3**	**4**	**3**	**3**	0	**3**	**3**	**4**
	Ranks 1–3	9	4	**13**	9	**11**	4	4	9	**10**
	Mean rank (sd)	3.3 (1.6)	5.3 (2.6)	**2.7 (1.4)**	3.3 (1.7)	3.9 (2.2)	3.5 (1.6)	4.1 (2.1)	**2.8 (1.6)**	**2.7 (1.5)**
	Conv. Prob's	**9**	**10**	**10**	13	**6**	22	15	16	13

Note: For notations, see Table 3.

Table 5. The best plot-specific fits of the 2- and 3- parameter models and the related RMSE in different datasets.

	Model name and RMSE (m)			
Dataset	2-parameter model		3-parameter model	
Scots pine A	Curtis	1.39	***Logistic***	***1.37***
Norway spruce A	Näslund	1.62	***Prodan***	***1.60***
Scots pine B	***Näslund***	***1.64***	Prodan	1.64
Norway spruce B	Näslund	1.27	***Prodan***	***1.21***
Birch	Näslund	1.97	***Logistic***	***1.92***
Norway spruce C	Näslund	2.01	***Gomperz***	***1.95***
Turkish red pine	Wykoff	1.95	***Prodan***	***1.95***
Canarian island pine	***Curtis***	***1.99***	Logistic	1.94
Aleppo pine	Näslund	0.97	***Prodan***	***0.97***
Loblolly pine 1	***Näslund***	***0.82***	Gomperz	0.82
Loblolly pine 2	***Schumacher***	***0.97***	Gomperz	0.97
Loblolly pine 3	**Schumacher**	**1.14**	*Prodan*	*1.13*
Lodgepole pine 1	Curtis	0.61	***Ratkowsky***	***0.61***
Lodgepole pine 2	Curtis	0.68	***Ratkowsky***	***0.68***
Lodgepole pine 3	Schumacher	0.85	***Ratkowsky***	***0.85***
Eugalyptus clone	***Schumacher***	***0.90***	Prodan	0.83
Blue gum A	Näslund	1.17	***Prodan***	***1.16***
Blue gum B1	***Näslund***	***2.27***	Ratkowsky	2.26
Blue gum B2	Näslund	2.29	***Logistic***	***2.20***
Centrolobium 1	**Curtis**	**1.26**	*Prodan*	*1.25*
Centrolobium 2	***Schumacher***	***1.28***	Ratkowsky	1.25
Centrolobium 3	***Schumacher***	***1.42***	Ratkowsky	1.39
Brasilian firetree	**Curtis**	**1.54**	*Prodan*	*1.53*
Teak 1	***Curtis***	***1.10***	Logistic	1.08
Teak 2	**Näslund**	**1.10**	*Gomperz*	*1.08*
Mixed tropical	Näslund	2.81	***Logistic***	***2.79***
Balsa 1	**Schumacher**	**1.23**	*Ratkowsky*	*1.23*
Balsa 2	**Schumacher**	**1.24**	*Prodan*	*1.24*

Note: The model with lower BIC value between the 2- and 3- parameter models is indicated by **boldface** and the model with lower AIC by *italics*.

Logistic model for Scots pine data

The previous comparisons of simple mixed-effects model for the pure Scots pine dataset (Scots pine A) suggested the logistic function for the plot-specific H–D relationship. The evaluation was made by having a random effect associated with each of the 3 parameters.

We first evaluated whether a model allowing heteroscedastic error variance improved the model fit. The class-specific standard deviations (Fig. 1, top) show some decrease in error variance as a function of relative tree diameter (see eq. (4)), but this trend was satisfactorily modeled by the power-type variance function where the predictor was the truncated relative diameter (Fig. 1, bottom).

Next, we explored whether the fit would get significantly worse if the number of random effects were reduced to two (for parameters a and b) or one (for a only). These restrictions increased the RMSE of predicted height appreciably, from 1.365 to 1.394 and 1.569 meters, respectively. Furthermore, the confidence intervals of class means in the middle and especially in the top graph of Fig. 2 showed a nonlinear trend, thus indicating lack of fit for the restricted model. For these reasons, all three random effects were retained in the model. These choices on variance function and random effects were supported also by an approximate LR test on the REML fits.

Next, we explored the correlation of the random effects with the plot-specific mean diameter and stand density. All three random effects had showed linear correlation with the untransformed mean diameter of the plot (Fig. 3), and including them significantly improved the model fit when evaluated using the conditional F-test. Therefore, the generalized mixed-effects model for the Scots pine A dataset was,

$$(7)\qquad h_{ij} = 1.3 + \frac{(\alpha_0 + \alpha_1\bar{d}_i + a_i)}{1 + (\beta_0 + \beta_1\bar{d}_i + b_i)\exp[(\gamma_0 + \gamma_1\bar{d}_i + c_i)d_{ij}]} + e_{ij}$$

where Greek letters are used for fixed coefficients, $\bar{d}_i$ is the mean diameter of plot i, and $\mathbf{b}_i = (a_i, b_i, c_i)'$ are the random effects, with trivariate normal distribution. The errors e_{ij} are independent, with mean of 0 and power-type variance function specified by eqs. (3) and (4). The parameter estimates of the model are shown in Table 6

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

Fig. 1. Residual plots of the logistic model in dataset Scots Pine A without (top) and with (bottom) a variance function. For notation, see Fig. 2.

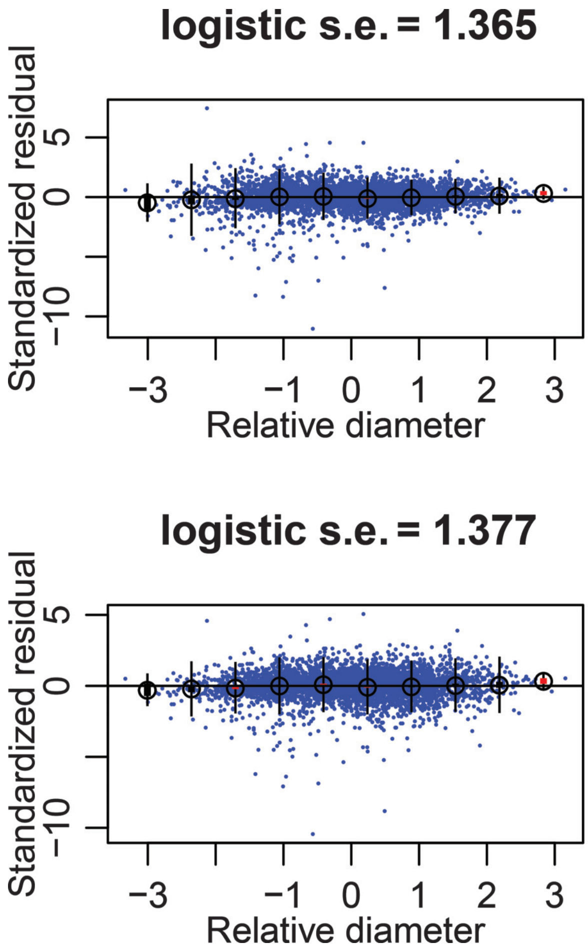
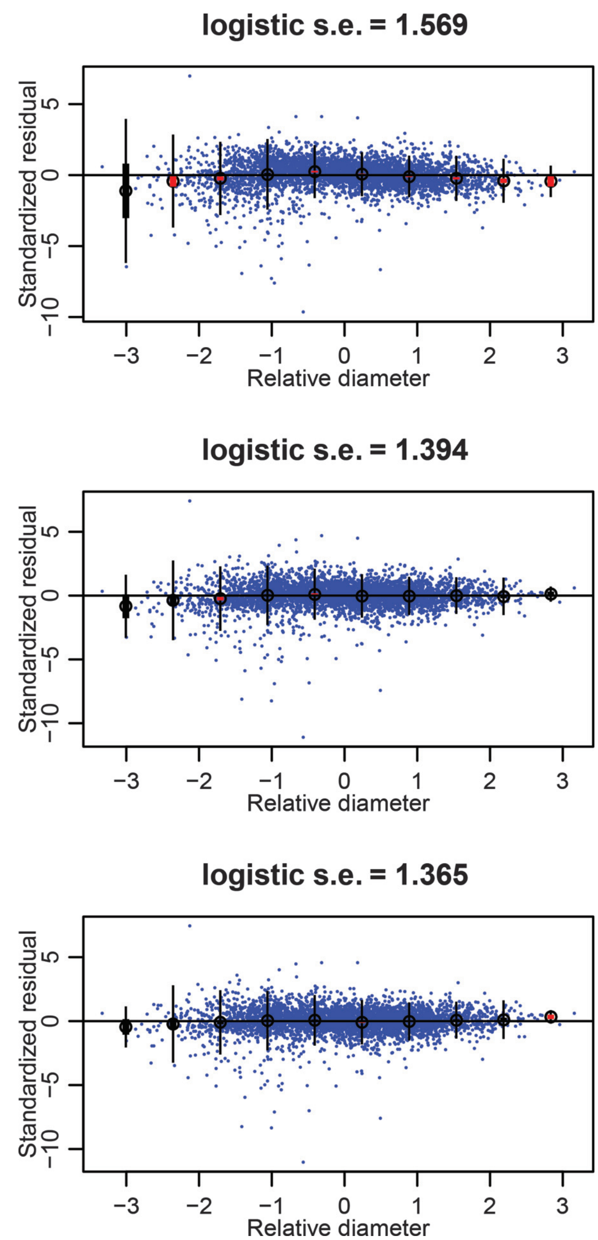


Fig. 2. Residual plots of the logistic model in dataset Scots Pine A when random effects were assigned for parameters *a* (top), *a* and *b* (middle), and *a*, *b* and *c* (bottom). The small dots show the residuals, and the large dots show the means of residuals in 10 relative diameter classes. The thin vertical lines show the confidence interval of one observation (mean ± 1.96 SD), and the thick vertical lines show the 95% confidence interval of class mean. The thick lines that do not cross the horizontal line at $y = 0$ are highlighted using red colour.



Näslund's model for young Loblolly pine data

The other datasets had similar steps for the evaluation of the variance function, need for random effects, and correlation of random effects with plot-specific predictors. To avoid unnecessary repetition, the model building procedure is not presented in detail, with a focus on the parts that show difference to the previously presented models.

For the first measurement occasion of the Loblolly pine dataset (dataset Loblolly pine 1), the best-fitting simple mixed-effects model was based on Näslund's function, which formed the basis for the generalized mixed-effects model. Random effects were needed for both parameters, and they showed linear dependence on logarithmic mean diameter. The final model was,

$$h_{ij} = 1.3 + \frac{d_{ij}^2}{\left[(\alpha_0 + \alpha_1 \ln \bar{d}_i + a_i)d_{ij} + \beta_0 + b_i + \beta_1 \ln \bar{d}_i\right]^2} + e_{ij} \tag{8}$$

where the decreasing error variance was modeled by the power function (3) by using the truncated relative diameter (eq. (4)) as the predictor. Table 7 shows the final parameter estimates of the model.

Curtis's model for Teak data

For the dataset Teak 1, the best-fitting simple mixed-effects model was Curtis's model. Similar to the previous study cases, we found linear relationship between the plot-specific mean diameter and all random effects, which led to the following final model,

$$h_{ij} = 1.3 + \frac{(\alpha_0 + \alpha_1 \bar{d}_i + a_i)d_{ij}}{(1 + d_{ij})^{\beta_0 + \beta_1 \bar{d}_i + b_i}} + e_{ij} \tag{9}$$

In contrast to previous models, the error variance increased as a function of tree diameter, but this trend was properly modeled by the power-type variance function as a function of tree diameter. Table 8 shows the final parameter estimates of the model.

Schumacher's model for Centrolobium data

The best-fitting model for the dataset Centrolobium 3 was Schumacher's function. The plot-specific mean diameter was corre-

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

Fig. 3. The random effects of the logistic simple mixed-effects model on the plot-specific mean diameter in the dataset Scots Pine A.

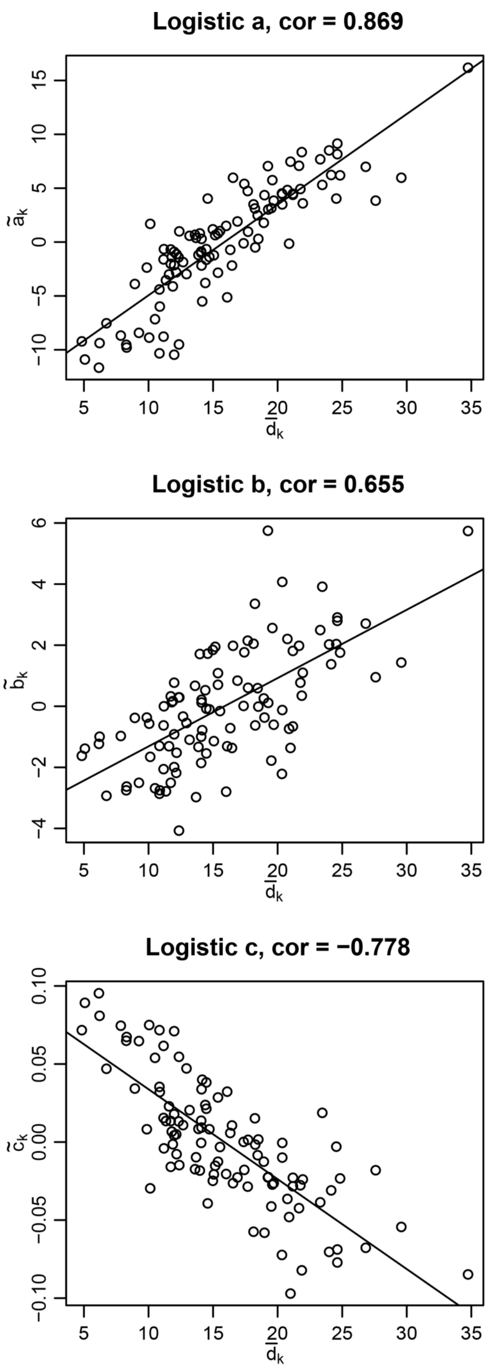


Table 6. The parameter estimates and standard errors of model (7) in the Scots pine A dataset.

Parameter	Estimate	SE
α_0	3.164	0.833
α_1	0.840	0.049
β_0	1.776	0.726
β_1	0.228	0.049
γ_0	0.291	0.016
γ_1	−0.005	0.001
sd (a_i)	2.681	
sd (b_i)	1.513	
sd (c_i)	0.0316	
corr (a_i, b_i)	0.281	
corr (a_i, c_i)	−0.668	
corr (b_i, c_i)	0.162	
σ^2	2.369^2	
δ	−0.542	

Table 7. The parameter estimates and standard errors of model (8) to the Loblolly pine 1 dataset.

Parameter	Estimate	SE
α_0	0.603	0.034
α_1	−0.133	0.013
β_0	2.350	0.310
β_1	−0.504	0.117
sd (a_i)	0.020	
sd (b_i)	0.171	
corr (a_i, b_i)	−0.151	
σ^2	1.002^2	
δ	−0.178	

Table 8. The parameter estimates and standard errors of model (9) in the Teak 1 dataset.

Parameter	Estimate	SE
α_0	10.549	0.686
α_1	0.616	0.079
β_0	6.632	0.287
β_1	0.0275	0.036
sd (a_i)	2.639	
sd (b_i)	1.021	
corr (a_i, b_i)	0.869	
σ^2	0.295^2	
δ	0.677	

lated with the random effect related to parameter a in the simple model, whereas the random effect associated with parameter b did not have significant correlation with mean diameter. This conclusion was based on Fig. 4 and was also further supported by an approximate conditional F-test (Pinheiro and Bates 2000). The decreasing error variance was modeled by using the power function with truncated relative diameter as the predictor. Therefore, the final model was,

$$h_{ij} = 1.3 + (\alpha_0 + \alpha_1\bar{d}_i + a_i)\exp[(\beta_0 + b_i)d_{ij}^{-1}] + e_{ij} \quad (10)$$

The parameter estimates of the model are shown in Table 9.

Model prediction in selected datasets

The four different models (simple and generalized mixed-effects and fixed-effects model) were used to compute six different predictions for the four datasets. The plot-specific predictions considered were (*i*) fixed-effect prediction from the simple mixed-effects model, (*ii*) fixed-effect prediction from the generalized mixed-effects model, (*iii*) random-effect prediction from the simple mixed-effects model, and (*iv*) random-effect prediction from the generalized mixed-effects model. The marginal predictions considered were the (*v*) prediction from the simple fixed-effects model and (*vi*) the prediction from the generalized fixed-effects model. The models used in these predictions were based on the best-fitting plot-specific function of each dataset, i.e., the same functions that were used earlier as the starting point in the generalized models. However, because the marginal H–D relationship in the data can be of different mathematical form than the plot-specific relationship, the simple fixed effects models was based on

Fig. 4. The correlation of the random effects related to parameters *a* and *b* of the simple Schumacher mixed-effects model with mean diameter in dataset Centrolobium 3.

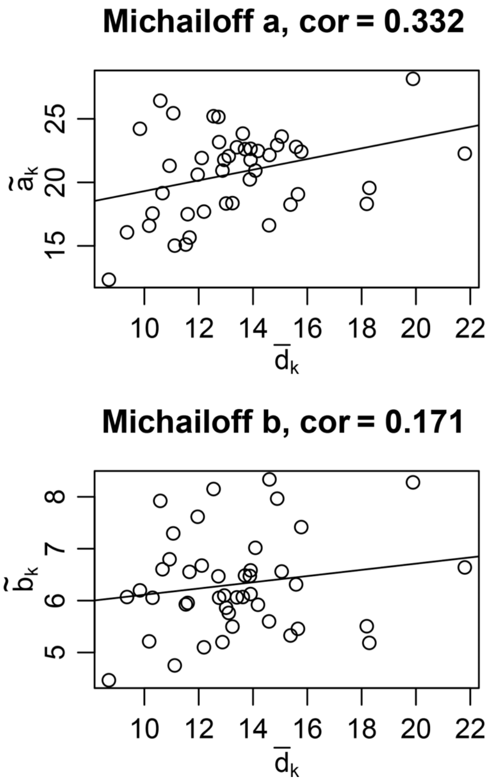


Table 9. The parameter estimates and standard errors of model (10) in the Centrolobium 3 dataset.

Parameter	Estimate	SE
α_0	15.560	1.975
α_1	0.360	0.142
β_0	6.059	0.210
sd (a_i)	3.256	
sd (b_i)	1.034	
corr (a_i,b_i)	0.677	
σ^2	2.475^2	
δ	−0.554	

the function that was found best-fitting for the marginal relationship as follows: Meyer's function for Scots pine A data, power function for Loblolly pine 1 data, Näslund's function for Centrolobium 3 data, and Prodan's model for Teak 1 data.

Table 10 shows the bias and RMSE of these predictions in the four datasets, and Figs. 5 and 6 illustrate the predictions in Scots pine A and Loblolly pine 1 datasets. When one considers the four fixed-effect predictions, the generalized models perform better than the corresponding simple model, because part of the between-plot variability is now explained by the plot-specific mean diameters. In addition, the marginal predictions from the fixed-effects models are better than the fixed-effect predictions from the mixed-effects models, when evaluated using RMSE and bias. This difference is sometimes used to support the use of fixed-effects model for prediction in hierarchical datasets. However, it is an expected result: the OLS fitting criteria in fixed-effects models is itself the mean-squared error. Therefore, it is not possible to find a model that would provide better fit than the fixed-effect model in the model fitting dataset, if OLS is used in model fitting and RMSE as the evaluation criterion. If a variance function is used, then the criterion is just a weighted sum of squares, and if random effects are included, then the criterion is a generalized least-squares criterion. In both cases, the predictions have necessarily higher RMSE than in the case of OLS fit for a linear model. For a nonlinear model, the RMSE of a mixed-effects model may be slightly lower than for a fixed-effects model.

The thick and moderately thick lines in Figs. 5 and 6 illustrate the four different fixed-effect predictions. In the case of simple fixed-effect predictions (the thickest lines), the fixed-effects models provide very different curves than the mixed-effects model. Also for generalized models, the fixed-effects model predictions are of different shape than the plot-specific relationship, even though these differences are not so striking as in the case of the simple models. Therefore, the fixed-effect models provide an H–D relationship that may be unrealistic when used as a plot-specific relationship, e.g., in a growth and yield simulator or tree height prediction in a forest inventory. Correspondingly, the fixed-effects prediction from the improperly formulated mixed-effects model cannot be interpreted as a marginal prediction. To conclude, one should not look only at the RMSE of the different predictions but rather recognize the interpretations of different predictions and use the one that serves best the purpose of model use.

The random-effect predictions are generally the best among the six predictions using the RMSE and bias criteria. The high accuracy is because all measured tree heights of the plot are directly utilized in the prediction of random plot effects. Therefore, the predictions take into account also that part of the between-plot variation that is not explained by the plot-specific mean diameters. In addition, the two mixed-effect predictions are equally accurate. This is also logical, because major proportion of the used information comes from the measured tree heights. However, if the number of trees per plot were low, then the fixed part would contribute more to the prediction, and the generalized mixed-effects model would give more accurate random-effect predictions than the simple mixed-effects model. Furthermore, decreasing the number of trees per plot would increase the RMSE towards the RMSE of the fixed-effect prediction of the mixed-effects model.

Discussion

This article has performed an extensive comparison of 16 nonlinear H–D models in 28 data sets from various geographical regions, ecological zones, and tree species. Because the aim was to compare the fit of different functions for plot-specific relationships, simple mixed-effects model without plot-specific predictors was sufficient at this stage; this also allowed automatic model selection, which was necessary due to the large number of datasets and models. About half of the BIC comparisons and one third of the AIC comparisons suggested a 2-parameter model, but the difference in accuracy of prediction was only slight in most cases. A 3-parameter model was suggested more often in the largest datasets of our study. However, the 3-parameter models sometimes suffered from problems in model convergence and nonpositive definite random-effect variance–covariance matrix. Similar problems of lack of convergence with increasing complexity of nonlinear H–D models have been also reported in previous research (e.g., Feldpausch et al. 2011). Therefore, we conclude that a 2-parameter model might be sufficient for most situations.

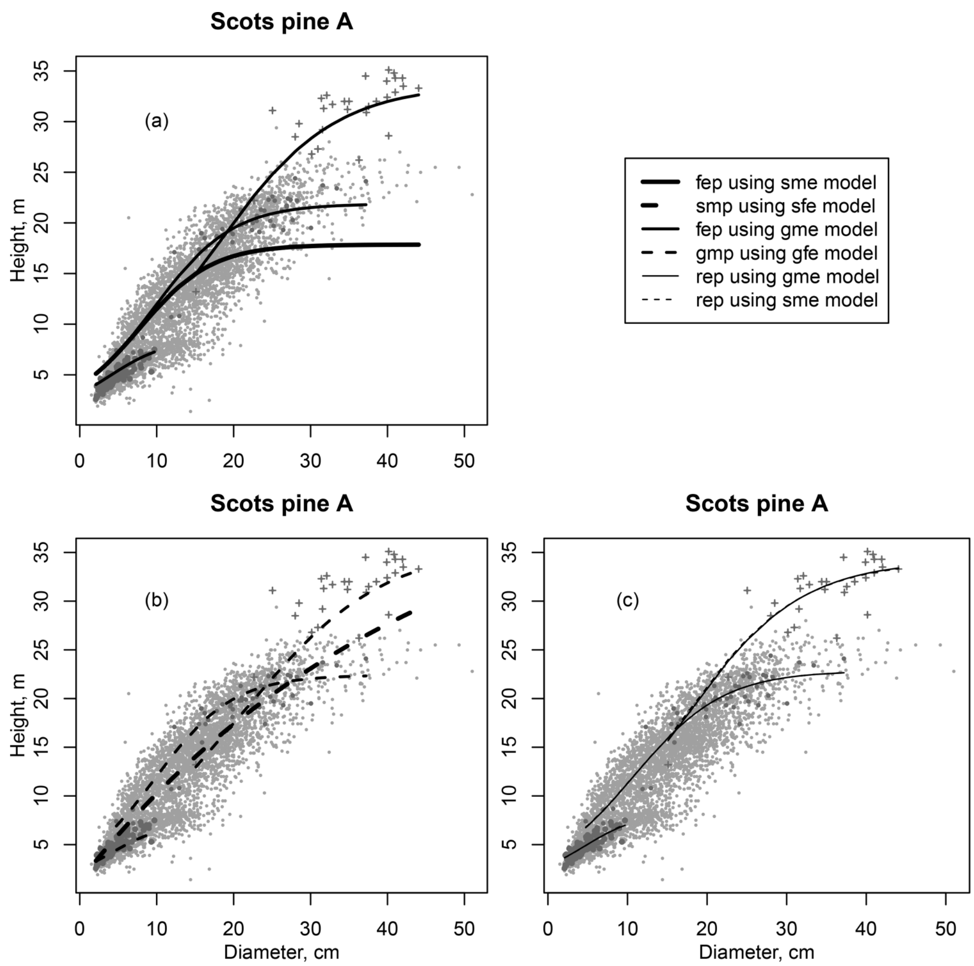
The Curtis, Näslund, and Schumacher functions were the best 2-parameter models for plot-specific relationship, and the models of Prodan and Gomperz were the best 3-parameter models. However, the selection of the function for a dataset should always be supported by graphical evaluation of model fit, e.g., by using similar graphs to those used in the detailed modeling part of this study.

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

Table 10. Empirical prediction bias and RMSE in selected modeling datasets using both simple and generalized fixed-effects and mixed-effects models and both fixed-effect and random-effect predictions.

Prediction	Plot-specific/Fixed-effect		Plot-specific/Random-effect		Marginal/Fixed-effect	
Model	Simple ME	Generalized ME	Simple ME	Generalized ME	Simple FE	Generalized FE
Bias (m)						
Scots pine A	–0.08	–0.09	–0.023	–0.023	0	0.01
Loblolly pine 1	0.559	0.228	–0.002	–0.002	0	0.002
Teak 1	0.187	0.013	0.013	0.011	0.002	0.022
Centrolobium 3	–0.045	–0.126	–0.012	0.009	0.003	–0.009
RMSE (m)						
Scots pine A	3.4	2.27	1.38	1.38	2.73	2.25
Loblolly pine 1	2.4	1.53	0.83	0.83	2.01	1.51
Teak 1	1.84	1.28	1.11	1.11	1.29	1.27
Centrolobium 3	2.16	2.09	1.42	1.43	2.1	2.07

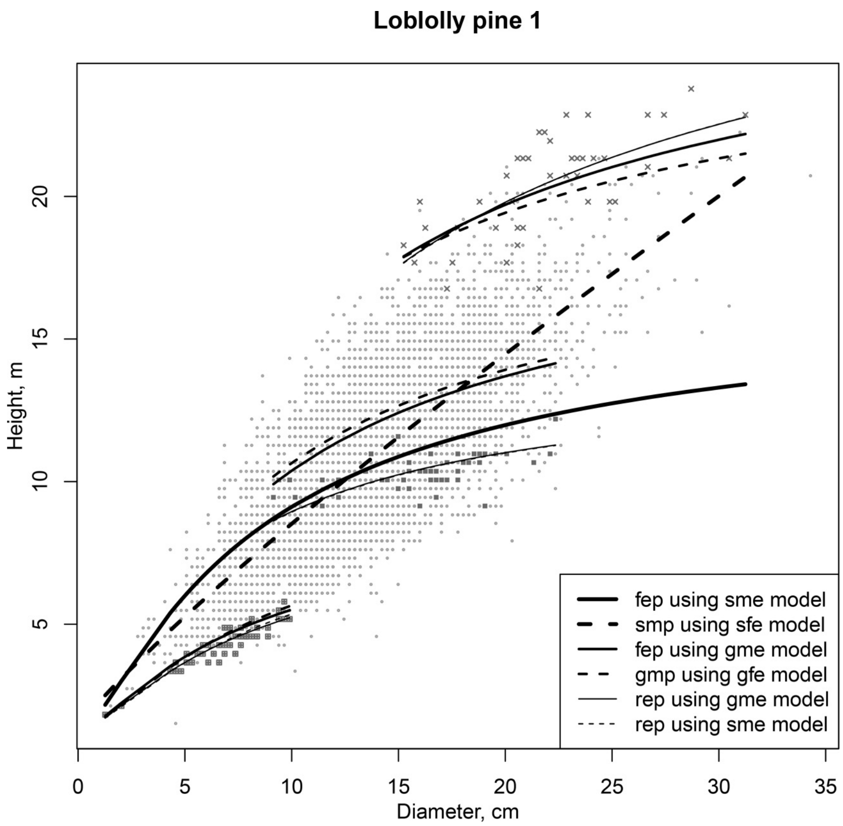
Fig. 5. Demonstration of different predictions in the Scots pine A dataset. The gray dots show the original data and the dark dots show the observed heights for 3 sample plots using plot-specific symbols. The figures show (*a*) the fixed-effect predictions (fep) using simple mixed-effects models (sme) and generalized mixed-effects models (gme) for the 3 sample plots (the predictions can be matched with the observations by the common range of diameters); (*b*) the marginal predictions (smp and gmp) using the simple and generalized fixed-effects models (sfe and gfe); and (*c*) the random-effect predictions (rep) from simple and generalized mixed-effects models.



The simple models were further explored for the correlation of plot-specific mean diameter with the random effects of the model in four example datasets. In all these datasets, significant correlation was detected between at least one of the random effects and the mean diameter. Therefore, the assumption of independent random effects was not met in the simple mixed-effects models, and a generalized model was used to satisfactorily meet the assumption of independent and identically distributed random effects. Moreover, a generalized model is needed whenever the model is used for inference, for prediction on plots with no calibration measurements, or for simulation of tree heights. However, the difference between the generalized and simple mixed-effects model is unimportant if moderately large number of observed H–D pairs from the

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

Fig. 6. Different predictions in Loblolly pine 1 dataset. Similar predictions are presented in Fig. 5 for Scots pine A dataset, but all predictions are pooled into a single figure to help in comparison.



Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.

sample plots of interest are used for plot-specific prediction using random effects.

No further suggestions on the model building can be given, based on our analysis. For example, some datasets were better modeled with the mean diameter as a predictor in both parameters of the generalized model, whereas others fitted well with mean diameter only in one parameter. Similar results have been reported also in previous papers: for example Feldpausch et al. (2011) used basal area as the predictor only in the intercept term of the model, whereas others have used plot-specific characteristics in several parameters (e.g., Soares and Tomé 2002; Lappi 1997). Furthermore, some datasets provided better fit by using logarithmic mean diameter, whereas others provided better fit with raw mean diameter. Finally, all but one dataset showed significantly inconstant variance, which was satisfactorily modelled using the power-type variance function. For some datasets, the tree diameter was a good predictor for the variance function, whereas in other datasets the relative tree diameter served the purpose better.

The parametric variance function improved the model fit in most datasets. It is quite common to model the error variance as a function of tree diameter or predicted height. Using this approach, the variance is usually found to increase as a function of diameter or predicted height. However in our detailed analyses, the truncated relative diameter (see eq. (4)) was a better predictor of the variance function than the raw diameter in three datasets out of four. The truncated relative diameter as a predictor in the power function assumes either increasing or decreasing variance for most trees of the plot and a constant variance for the very smallest trees. In those three datasets where this function was the best, the power parameter of the function was negative, implying decreasing variance as a function of tree diameter within the plot. This phenomenon can be explained by the biological growth rhythm of (especially shade-intolerant) trees in dense forests: they first concentrate on height growth to compete on light and grow in diameter only after surviving this competition. This results in higher height variation for small-diameter trees than for large-diameter trees. This phenomenon is clearly visible when residuals are plotted on relative tree diameter. If the residual plots were constructed based on raw diameter, then different sample plots would at least partially cancel the effects of each other, because trees of a certain diameter would be the smallest trees in one stand and thickest in another. Furthermore, stands with higher mean diameter tended to result in more slender trees (i.e., taller trees at any given tree diameter), which may be also explained in terms of greater symmetric competition for light in stands with high mean diameter (correlated with high stand basal area). Such competition would drive the allocation of more resources to height growth versus diameter increment (Feldpausch et al. 2011).

If the fitted H–D models are used in stochastic simulations, one might additionally wish to add a random residual to the random-effect prediction to make the properties of the imputed tree data as similar to the measured data as possible. Such a simulation was also implemented in the functions of package lmfor.

Figures 5 and 6 illustrate that the marginal and plot-specific H–D curves for a given data may be very different in shape and are usually best described with different mathematical functions. From the modeler's point of view, this implies that a good marginal fit of a certain function should not be taken as a support for the same function at plot level. For example, Meyer's function was one of the best functions for simple marginal H–D relatonship in our evaluation of simple 2-parameter models (Table 3), but one of the worst functions for plot-specific relationship. Therefore, it is important to make the distinction between plot-specific and different marginal predictions. Otherwise, straightforward comparison of the model using widely used RMSE and bias as criteria would lead to favouring fixed-effects models and OLS fitting criteria in all cases. Notably, conditional on the assumed model shape, the least-squares estimates will always provide the lowest mean-squared error in the modeling dataset; in separate evaluation datasets the situation may not be as straightforward (e.g., de-Miguel et al. 2012, 2013). However, the RMSE criterion itself may not be enough and can be even misleading. It may provide a shape for the predicted H–D relationship that is very unlikely on a single plot, in addition to the problems of inference caused by ignoring the dependence of observations from the same plot.

Acknowledgements

We thank Matti Maltamo, Timo Pukkala, Forest Modeling Research Cooperative at Virginia Tech, British Columbia Forest Service, and the Universidad Mayor de San Simón for providing their datasets for use in this paper.

References

Arabazis, A.A., and Burkhart, H.E. 1992. An evaluation of sampling methods and model forms for estimating height-diameter relationships in loblolly pine plantations. For. Sci. **38**: 192–198.

Box, M.J. 1971. Bias in nonlinear estimation. J. R. Stat. Soc. Series B Stat. Methodol. **33**(2): 171–201.

Calama, R., and Montero, G. 2004. Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. Can. J. For. Res. **34**: 150–163. doi:10.1139/x03-199.

Curtis, R.O. 1967. Height-diameter and height-diameter age equations for second-growth Douglas fir. For. Sci. **13**(4): 365–375.

de-Miguel, S., Guzman, G., and Pukkala, T. 2013. A comparison of fixed- and mixed-effects modeling in tree growth and yield prediction of an indigenous neotropical species (*Centrolobium tomentosum*) in a plantation system. For. Ecol. Manage. **291**: 249–258. doi:10.1016/j.foreco.2012.11.026.

de-Miguel, S., Mehtätalo, L., Shater, Z., Kraid, B., and Pukkala, T. 2012. Evaluating marginal and conditional predictions of taper models in the absence of calibration data. Can. J. For. Res. **42**(7): 1383–1394. doi:10.1139/x2012-090.

Feldpausch, T.R., Banin, L., Phillips, O.L., Baker, T.R., Lewis, S.L., Quesada, C.A., Affum-Baffoe, K., Arets, E.J.M.M., Berry, N.J., Bird, M., Brondizio, E.S., de Camargo, P., Chave, J., Djagbletey, G., Domingues, T.F., Drescher, M., Fearnside, P.M., França, M.B., Fyllas, N.M., Lopez-Gonzalez, G., Hladik, A., Higuchi, N., Hunter, M.O., Iida, Y., Salim, K.A., Kassim, A.R., Keller, M., Kemp, J., King, D.A., Lovett, J.C., Marimon, B.S., Marimon-Junior, B.H., Lenza, E., Marshall, A.R., Metcalfe, D.J., Mitchard, E.T.A., Moran, E.F., Nelson, B.W., Nilus, R., Nogueira, E.M., Palace, M., Patiño, S., Peh, K.S.H., Raventos, M.T., Reitsma, J.M., Saiz, G., Schrodt, F., Sonké, B., Taedoumg, H.E., Tan, S., White, L., Wöll, H., and Lloyd, J. 2011. Height-diameter allometry of tropical forest trees. Biogeosciences, **8**(5): 1081–1106. doi:10.5194/bg-8-1081-2011.

Flewelling, J.W., and de Jong, R. 1994. Considerations in simultaneous curve fitting for repeated height-diameter measurements. Can. J. For. Res. **24**: 1408–1414. doi:10.1139/x94-181.

Garber, S.M., Temesgen, H., Monleon, V.J., and Hann, D.W. 2009. Effects of height imputation strategies on stand volume estimation. Can. J. For. Res. **39**(3): 681–690. doi:10.1139/X08-188.

Gomperz, B. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philos. Trans. R. Soc. Lond. B Biol. Sci. **115**: 513–585.

Houghton, D.R., and Gregoire, T.G. 1993. Minimum subsamples of tree heights for accurate estimation of Loblolly pine volume. South. J. Appl. For. **17**(3): 124–129.

Huang, S., Price, D., and Titus, S.J. 2000. Development of ecoregion-based height-diameter models for white spruce in boreal forests. For. Ecol. Manage. **129**: 125–141. doi:10.1016/S0378-1127(99)00151-6.

Huang, S., Titus, S.J., and Wiens, D.P. 1992. Comparison of nonlinear height–diameter functions for major Alberta tree species. Can. J. For. Res. **22**: 1297–1304. doi:10.1139/x92-172.

Lappi, J. 1986. Mixed linear models for analyzing and predicting stem form variation of Scots pine. Communicationes Instituti Forestalis Fenniae 134, FFRI. 69 p.

Lappi, J. 1991. Calibration of height and volume equations with random parameters. For. Sci. **37**: 781–801.

Lappi, J. 1997. A longitudinal analysis of height/diameter curves. For. Sci. **43**: 555–570.

Lappi, J., and Bailey, R. 1988. A height prediction model with random stand and tree parameters: an alternative to traditional site index methods. For. Sci. **34**: 907–927.

Lundqvist, B. 1957. On the height growth in cultivated stands of pine and spruce in Northern Sweden. Medd. frstatens skogforsk 133.

Lynch, T.B., Holley, A.G., and Stevenson, D.J. 2005. A random-parameter height-dbh model for cherrybark oak. South. J. Appl. For. **29**(1): 22–26.

Mehtätalo, L. 2004. A longitudinal height-diameter model for Norway spruce in Finland. Can. J. For. Res. **34**(1): 131–140. doi:10.1139/x03-207.

Mehtätalo, L. 2005. Height-diameter models for Scots pine and birch in Finland. Silva Fenn. **39**(1): 55–66.

Mehtätalo, L. 2015. lmfor: Functions for Forest Biometrics. Available from http://cran.r-project.org/web/packages/lmfor/index.html.

Meng, S.X., and Huang, S. 2009. Improved calibration of nonlinear mixed-effects models demonstrated on a height growth function. For. Sci. **55**(3): 238–248.

Menten, L., and Michaelis, M.I. 1913. Die Kinetik der Invertinwirkung. Biochem. Z. **49**: 333–369.

Meyer, W. 1940. A mathematical expression for height curves. J. For. **38**: 415–420.

Michailoff, I. 1943. Zahlenmässiges verfahren für die ausführung der bestand-eshhenkurven. Forstw. Clb U Thar. Forstl. Jahrb. **6**: 273–279.

Moore, J.A., Zhang, L., and Stuck, D. 1996. Height-diameter equations for ten tree species in the inland northwest. West. J. Appl. For. **11**: 132–137.

Näslund, M. 1937. Skogsförsöksanstaltens gallringsförsök i tallskog (Forest research intitute's thinning experiments in Scots pine forests). Meddelanden frstatens skogsförsöksanstalt Häfte 29 (In Swedish).

Paulo, J.A., Tomé, J., and Tomé, M. 2011. Nonlinear fixed and random generalized height-diameter models for Portuguese cork oak stands. Ann. For. Sci. **68**: 295–309. doi:10.1007/s13595-011-0041-y.

Pearl, R., and Reed, L.J. 1920. On the rate of growth of the population of the united states since 1790 and its mathematical representation. Proc. Natl. Acad. Sci. U.S.A. **6**: 275–288. doi:10.1073/pnas.6.6.275.

Peschel, W. 1938. Mathematical methods for growth studies of trees and forest stands and the results of their application. Tharandter Forstl. Jahrb. **89**: 169–247.

Pinheiro, J.C., and Bates, D.M. 2000. Mixed-effects models in S and Splus. Springer-Verlag, New York, U.S.A. 528 p.

R Development Core Team. 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org. ISBN 3-900051-07-0.

Ratkowsky, D.A. 1990. Handbook of nonlinear regression. Marcel Deccer Inc., New York, U.S.A.

Richards, F.J. 1959. A flexible growth function for empirical use. J. Exp. Bot. **10**(29): 290–300. doi:10.1093/jxb/10.2.290.

Schumacher, F.X. 1939. A new growth curve and its application to timber yield studies. J. For. **37**: 819–820.

Sibbesen, E. 1981. Some new equations to describe phosphate sorption by soils. J. Soil Sci. **32**: 67–74. doi:10.1111/j.1365-2389.1981.tb01686.x.

Siipilehto, J. 1999. Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fenn. **33**: 281–301. doi:10.14214/sf.650.

Sirkiä, S., Heinonen, J., Miina, J., and Eerikäinen, K. 2014. Subject-specific prediction using a nonlinear mixed model: consequences of different approaches. For. Sci. In press. doi:10.5849/forsci.13-142.

Soares, P., and Tomé, M. 2002. Height-diameter equation for first rotation eucalypt plantation in Portugal. For. Ecol. Manage. **166**: 99–109. doi:10.1016/S0378-1127(01)00674-0.

Stoffels, A., and van Soest, J. 1953. The main problems in sample plots. Ned. Boschb. Tijdschr. **25**: 190–199.

Strand, L. 1959. The accuracy of some methods for estimating volume and increment on sample plots. Medd. Norske Skogfors. **15**(4): 284–392 (In Norwegian with English summary).

Temesgen, H., and von Gadow, K. 2004. Generalized height-diameter models – an application for major tree species in complex stands of iterior British Columbia. Eur. J. For. Res. **123**: 45–51. doi:10.1007/s10342-004-0020-z.

Temesgen, H., Monleon, V.J., and Hann, D.W. 2008. Analysis and comparison of nonlinear tree height prediction strategies for douglas-fir forests. Can. J. For. Res. **38**(3): 553–565. doi:10.1139/X07-104.

Weibull, W. 1951. A statistical distribution function of wide applicability. J. Appl. Mech. Trans. ASME, **18**(3): 293–297.

Wykoff, W.R., Crookston, N.L., and Stage, A.R. 1982. User's guide to the stand prognosis model. USDA For. Serv. Gen. Tech. Rep. 133.

Zeide, B. 1993. Analysis of growth equations. For. Sci. **39**(3): 594–616.

Can. J. For. Res. Downloaded from www.nrcresearchpress.com by University of Eastern Finland on 08/17/15
For personal use only.